(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号
特開2002－14990
（P2002－14990A）
(43)公開日　平成14年１月18日(2002.1.18)

(51)Int.Cl.⁷	識別記号	ＦＩ	ﾃｰﾏｺｰﾄﾞ(参考)
Ｇ０６Ｆ　17/30	３３０	Ｇ０６Ｆ　17/30	３３０Ｃ　５Ｂ０７５
	１７０		１７０Ａ
	３２０		３２０Ｄ
9/44	５７０	9/44	５７０Ｃ

審査請求　有　　請求項の数７　ＯＬ　（全 10 頁）

(21)出願番号　特願2000－193671(P2000－193671)

(22)出願日　平成12年６月28日(2000.6.28)

(71)出願人　301022471
独立行政法人通信総合研究所
東京都小金井市貫井北町４－２－１

(72)発明者　村田　真樹
兵庫県神戸市西区岩岡町岩岡588－２　郵政省通信総合研究所　関西先端研究センター内

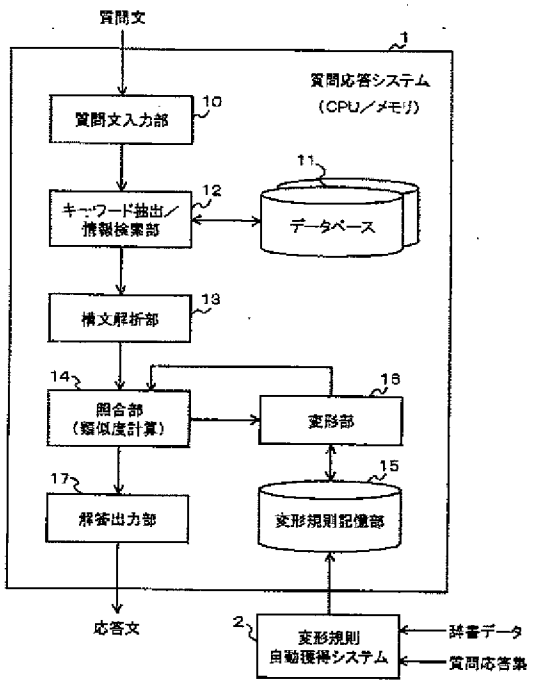
(72)発明者　内山　将夫
兵庫県神戸市西区岩岡町岩岡588－２　郵政省通信総合研究所　関西先端研究センター内

(74)代理人　100087848
弁理士　小笠原　吉義

最終頁に続く

(54)【発明の名称】　質問応答システム，質問応答処理方法，変形規則自動獲得処理方法およびそれらのプログラム記録媒体

(57)【要約】
【課題】　質問文に対する解答の正解率が高く，かつシステムの構築が容易で柔軟性のある質問応答システムを提供する。
【解決手段】　質問文に対する解答をデータベース11から得るために，照合部14において，質問文とデータベース文とを照合し類似度を計算するとともに，変形部16において，あらかじめ変形規則記憶部15に記憶された変形規則を用いて質問文とデータベース文とを変形し，照合部14による照合と変形部16による文の変形とを繰り返し，質問文との類似度が高くなるデータベース文を探索する。



【特許請求の範囲】

【請求項１】　自然言語による質問文を入力し，データベース中の文との照合によって応答文を生成して出力する質問応答システムにおいて，文を同じ内容を表す他の文に変形する規則を記憶する変形規則記憶部と，入力した質問文とデータベースから抽出した文とを照合し，それらの類似度を算出する照合部と，前記照合部による類似度の算出結果に基づき，前記変形規則記憶部に記憶されている変形規則を用いて前記質問文と前記データベースから抽出した文とを書き換える変形部と，前記照合部と前記変形部とによる処理を繰り返した後，前記類似度が最も高くなる照合において抽出された解を応答文として出力する解答出力部とを備えることを特徴とする質問応答システム。

【請求項２】　自然言語による質問文を入力し，データベース中の文との照合によって応答文を生成して出力する質問応答処理方法において，入力した質問文とデータベースから抽出した文とを照合し，それらの類似度を算出する過程と，あらかじめ記憶されている文の変形規則を用いて，前記質問文と前記データベースから抽出した文とを，それらの類似度が最も高くなるまで書き換える過程と，前記類似度が最も高くなる照合において抽出された解を応答文として出力する過程とを有することを特徴とする質問応答処理方法。

【請求項３】　自然言語で記述された文を同じ内容を表す他の文に変形する変形規則をコンピュータを用いて生成する方法であって，複数の辞書ファイルから読み出した辞書データから同じ単語の説明文を抽出する過程と，抽出した複数の説明文を突き合わせ，その結果から同義語または同義フレーズを抽出する過程と，抽出した同義語または同義フレーズから，ある文を同じ内容を表す他の文に書き換えるための変形規則を生成する過程とを有することを特徴とする変形規則自動獲得処理方法。

【請求項４】　自然言語で記述された文を同じ内容を表す他の文に変形する変形規則をコンピュータを用いて生成する方法であって，質問文とそれに対する応答文とを入力する過程と，入力した質問文と応答文とを突き合わせ，その結果から同義語または同義フレーズを抽出する過程と，抽出した同義語または同義フレーズから，ある文を同じ内容を表す他の文に書き換えるための変形規則を生成する過程とを有することを特徴とする変形規則自動獲得処理方法。

【請求項５】　自然言語による質問文を入力し，データベース中の文との照合によって応答文を生成して出力するためのプログラムを記録した記録媒体であって，入力した質問文とデータベースから抽出した文とを照合し，それらの類似度を算出する処理と，あらかじめ記憶されている文の変形規則を用いて，前記質問文と前記データベースから抽出した文とを，それらの類似度が最も高くなるまで書き換える処理と，前記類似度が最も高くなる照合において抽出された解を応答文として出力する処理とを，コンピュータに実行させるためのプログラムを記録したことを特徴とする質問応答処理プログラム記録媒体。

【請求項６】　自然言語で記述された文を同じ内容を表す他の文に変形する変形規則をコンピュータを用いて生成するためのプログラムを記録した記録媒体であって，複数の辞書ファイルから読み出した辞書データから同じ単語の説明文を抽出する処理と，抽出した複数の説明文を突き合わせ，その結果から同義語または同義フレーズを抽出する処理と，抽出した同義語または同義フレーズから，ある文を同じ内容を表す他の文に書き換えるための変形規則を生成する処理とを，コンピュータに実行させるためのプログラムを記録したことを特徴とする変形規則自動獲得処理プログラム記録媒体。

【請求項７】　自然言語で記述された文を同じ内容を表す他の文に変形する変形規則をコンピュータを用いて生成するためのプログラムを記録した記録媒体であって，質問文とそれに対する応答文とを入力する処理と，入力した質問文と応答文とを突き合わせ，その結果から同義語または同義フレーズを抽出する処理と，抽出した同義語または同義フレーズから，ある文を同じ内容を表す他の文に書き換えるための変形規則を生成する処理とを，コンピュータに実行させるためのプログラムを記録したことを特徴とする変形規則自動獲得処理プログラム記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は，コンピュータによる自然言語の情報処理システムに係わり，特に類似度に基づく推論を用いた質問応答システムに関するものである。情報検索，情報抽出に利用することができる。

【０００２】

【従来の技術】質問応答システムとは，例えば「パーキンソン病の兆候は脳のどの部分にある細胞の死が関係していますか。」という質問を入力すると，大量の電子化テキストから「パーキンソン病は，中脳の黒質にあるメラニン細胞が変性し，黒質細胞内で作られる神経伝達物質のドーパミンがなくなり発病する，とされている。」といった文を探し出し，その文の中から質問に該当する「黒質」を的確に取り出しこれを解答として出力するシステムのことである。

【０００３】このような質問応答システムとして，本発明者等が下記の参考文献１に発表した「構文情報を利用した質問応答システム」が知られている。

〔参考文献１〕Masaki Murata and Masao Utiyama and Hitoshi Isahara, Question Answering System Using Syntactic Information, 1999, http://xxx.lanl.gov/abs/cs.CL/9911006.

この質問応答システムでは，まず質問文からキーワード

抽出を行い，データベースからキーワードのＩＤＦ（Inverse Document Frequency）の和が大きい文を抽出する。次に質問文と抽出した文とを構文解析する。解析結果の構文情報を利用して質問文と抽出した文とを照合し，解の候補を出しつつ，質問文と抽出した文の類似度を所定の算出方法に従って計算する。データベースから抽出した文のうち質問文との類似の最も高かったものから解を抽出する。解の抽出は，質問文における疑問詞を含む文節に対応づけられたデータベース側の文節を解とすることで行う。このシステムは，質問文とデータベースから得た文を類似度が高くなるように変形して照合することは考慮していない。

【０００４】この「構文情報を利用した質問応答システム」に先行する従来技術として，下記の参考文献２に記載されている文の変形を利用する情報検索技術が知られている。以下，これをＫａｔｚの方法という。

〔参考文献２〕Boris Katz, Using English for Indexing and Retrieving, Artificial Inteiligence at MIT, Vol.2, MIT Press, 1990.

一般に，質問応答システムにおける質問文とデータベースの文の照合による解答の導出では，質問文とデータベース文とを照合することにより，質問文に最も一致するデータベース文の中から解答を得ることが行われる。

【０００５】例えば「日本の首都はどこですか」という入力質問文があったときには，それと良く似た文をデータベースから抽出する。ここでは，「日本の首都は東京である」という文があったとしよう。そして，「日本の首都はどこですか」と「日本の首都は東京である」とを照合し，疑問詞「どこ」の部分に対応する「東京」を解答として出力する。

【０００６】しかし，いつも上記のように質問文とデータベース文とがぴったり一致し，解答部分を容易に取り出せるとは限らない。例えば，データベースには，「日本の首都は東京である」という文がなく，「東京は日本の首都である」という文しかなかったとしよう。そうすると，質問文と照合できず解が得られなくなってしまう。

【０００７】上記参考文献２に記載されているＫａｔｚの方法は，質問文とデータベース文との文型が異なっても照合ができるように，すべての文を最も一般的な表現（これを基底表現という）に変形してから照合を行って解を抽出する。例えば，上記の例では「日本の首都は東京である」と「東京は日本の首都である」と，どちらが基底表現かはわからないが，ここでは「日本の首都は東京である」を基底表現としておくことにして，データベースの文をそれに変形し，質問文の「日本の首都はどこですか」と照合して解を得る。

【０００８】

【発明が解決しようとする課題】上記参考文献１の「構文情報を利用した質問応答システム」は，質問文とデータベース文との照合に主に意味制約を利用するものである。しかし，文を照合する際に意味制約を利用する方法と文の変形を利用する方法には，それぞれ一長一短があり，意味制約を利用する方法だけでは，必ずしも十分な照合を実現できるとは限らない。文の変形を利用する方法も質問文とデータベース文との照合には有効であると考えられる。

【０００９】しかし，上記Ｋａｔｚの方法には，次のような難点がある。それは，すべての文を最も一般的な表現である基底表現に変形する必要があるが，基底表現を厳密に定義することは困難であることである。例えば，基底表現を能動態の文と定義し，受動態の文を能動態の文に変形するような場合には，比較的一律な変形が可能であるが，上記の例で「東京は日本の首都である」という文と，「日本の首都は東京である」という文とは，どちらを基底表現とするかを明確に定めることはできない。すなわち，相互に変形可能な複数の文があった場合に，どちらを基底表現とすべきか曖昧なケースが数多くあり，必ず基底表現に変形するＫａｔｚの方法では，すべてのケースについてどちらかを無理やり基底表現と定義する必要がある。しかし，このような基底表現を決めるのは実際には非常に困難な作業である。

【００１０】また，単なる基底表現への変形では，「日本の首都は東京である」という文と，「東京は関東地方にある」という二つの文から「日本の首都は関東地方にある」というような文を導出する変形はできない。

【００１１】本発明は上記問題点の解決を図り，質問文に対する解答の正解率が高く，かつシステムの構築が容易で柔軟性のある質問応答システムを提供することを目的とする。具体的には，質問応答システムにおいて文の変形を利用するにあたって，基底表現を決める必要をなくし，変形規則の記述を容易に行うことができるようにすること，文の変形を柔軟に行うことができるようにすること，また変形規則を自動獲得する手段を提供することを目的とする。

【００１２】

【課題を解決するための手段】本発明は，質問文とデータベース文との照合の際に，あらかじめ記憶された変形規則を用いて，質問文とデータベース文との類似度が高まるように双方の文を書き換えることを最も主要な特徴とする。

【００１３】変形を利用するところは上記Ｋａｔｚの方法に似ているが，Ｋａｔｚの方法では，すべての文を基底表現に変形するのに対し，本発明では，基底表現への変形に限らず，類似度を尺度として，類似度が高くなるように変形を行う。これにより，本発明には以下の利点がある。

【００１４】本発明では，基底表現を決める必要性がないため，変形規則の記述が容易になる。例えば，「日本の首都は東京である」と「東京は日本の首都である」が

相互に変形可能な場合，どちらを基底表現とすべきか曖昧であり，必ず基底表現に変形するＫａｔｚの方法では，どちらかを無理やり基底表現と定義する必要がある。しかし，このように無理やり定義しなければならない場合は数多くあり，基底表現を決めるのは実際的に難しい。これに対して本発明では，必ずしも基底表現への変形である必要はないため，この例の場合には，以下の二つの規則を書くことで問題が解決される。

【００１５】規則１：「日本の首都は東京である」を「東京は日本の首都である」に変形
規則２：「東京は日本の首都である」を「日本の首都は東京である」に変形

このように本発明は，変形規則の右辺が基底表現である必要がないことが特徴である。ただし，右辺に基底表現以外のものを記述する場合，変形を制御・管理する機構が必要であり，本発明では類似度という尺度で変形を制御・管理する。つまり，類似度が高くなるように変形を行うことで，質問文とデータベース文との照合における精度の向上を図る。

【００１６】また，本発明は，基底表現への変形規則という制約がないため，国語辞典などの辞書データや既存の質問応答集などのデータを利用して変形規則を自動獲得することも可能である。すなわち，複数の辞書ファイルから読み出した辞書データから同じ単語の説明文（または定義文）を抽出し，抽出した複数の説明文を突き合わせ，その結果から同義語または同義フレーズを抽出して，ある文を同じ内容を表す他の文に書き換えるための変形規則を生成する。

【００１７】または，既にある質問文とそれに対する応答文とを入力し，入力した質問文と応答文とを突き合わせ，その結果から同義語または同義フレーズを抽出し，それをもとに変形規則を生成する。

【００１８】以上の各処理をコンピュータによって実現するためのプログラムは，コンピュータが読み取り可能な可搬媒体メモリ，半導体メモリ，ハードディスクなどの適当な記録媒体に格納することができる。

【００１９】

【発明の実施の形態】図１は，本発明のシステム構成例を示す。図中，１は本発明に係る質問応答システム，２は辞書データや質問応答集から文の変形規則を自動獲得する変形規則自動獲得システムである。

【００２０】質問文入力部１０は，自然言語による質問文を入力する手段である。データベース１１は，新聞，論文その他各種文献の電子化されたテキスト情報が格納されたデータベースである。キーワード抽出／情報検索部１２は，質問文からキーワードを抽出し，データベースを検索する手段である。構文解析部１３は，質問文とデータベース１１から検索によって抽出された文（これをデータベース文という）とを構文解析する手段である。照合部１４は，入力した質問文とデータベース文とを照合し，それらの類似度を算出する手段である。変形規則記憶部１５は，文を同じ内容を表す他の文に変形する規則を記憶しているものである。

【００２１】変形部１６は，変形規則記憶部１５に記憶されている変形規則を用いて質問文とデータベース文とを書き換える手段である。書き換えた結果は，再度，照合部１４において照合され，類似度が算出され，変形部１６による処理と照合部１４による処理とが，類似度が向上しなくなるまで繰り返される。解答出力部１７は，類似度が最も高くなる照合において抽出されたデータベース文から解を抽出し，それを応答文として出力する手段である。

【００２２】図２に，変形規則記憶部１５に格納される変形規則の例を示す。図２（Ａ）は，同義語についての変形規則の例であり，この変形規則は，上記Ｋａｔｚの方法でも扱えるものである。図２（Ｂ）は，同義フレーズについての変形規則の例であり，この例における「ＡはＢである」→「ＢはＡである」という変形規則は，上記Ｋａｔｚの方法では扱うことはできない。

【００２３】また，以上のような意味の直接的な等価性を扱うものだけではなく，推論に関与する変形規則を利用することもできる。図２（Ｃ）に，その例を示す。ここでは，変形規則の左辺が「Ａである」と「ＡならばＢである」という複数の文を入力としている。この利用例について説明する。

【００２４】例えば，「晴れである」という文と「晴れならば傘は不要である」という文の二つの文があったとする。これらの文に，図２（Ｃ）に示す変形規則を適用すると，「晴れ」とＡ，「傘は不要」とＢが一致し，その結果から「傘は不要である」が導出される。

【００２５】本発明では，このように推論によって得られる知識も変形規則で扱うことができる。変形規則の左辺，右辺には，どのようなものがきてもよく，文の一部でも結合体でも任意の記述が可能である。

【００２６】次に，図３に示すフローチャートに従って，図１に示す質問応答システム１の処理を説明する。

【００２７】まず，質問文入力部１０が質問文を入力する（ステップＳ１）。ネットワークを介した端末からの入力，または情報検索などのアプリケーションプログラムからの入力など，入力方法は問わない。

【００２８】キーワード抽出／情報検索部１２は，入力した質問文からキーワードを抽出する（ステップＳ２）。キーワード抽出の簡単な方法としては，例えば文を形態素解析し，名詞のみを残すといった方法がある。

【００２９】次に，キーワード抽出／情報検索部１２は，データベース１１からキーワードのＩＤＦ（Inverse Document Frequency）の和が大きい文を複数文抽出する（ステップＳ３）。ＩＤＦの値の簡単な算出方法としては，例えばデータベース中の全文字列をＮ，キーワードのデータベース中での出現頻度をｎとするときに，１

ｏｇ（Ｎ／ｎ）としたものを用いることができる。なお，ＩＤＦの値を用いずに，データベース中からおおざっぱに質問文に現れるキーワードを含む文をすべて取り出してもよい。

【００３０】次に，構文解析部１３は，質問文とデータベース１１から抽出した文（データベース文）のすべてを構文解析し，これらをそれぞれ質問文の集合，データベース文の集合とする（ステップＳ４）。この構文解析では，例えば次の参考文献３に記載されている日本語構文解析システムなどを利用することができる。

〔参考文献３〕Sadao Kurohashi,Japanese Dependency/Case Structure Analyzer KNP version 2.0b6,(Department of Informatics,Kyoto Univercity,1998).

その後，照合部１４は，現在までの最も大きい類似度を記憶する変数Ｓを０に初期化し（ステップＳ５），ステップＳ６に進む。ステップＳ６では，構文解析部１３による構文情報を利用して，質問文の集合の各成員と，データベース文の集合の各成員をあらゆる組合せで照合し，それぞれに対して解の候補を求めながら，質問文とデータベース文の類似度を計算する。

【００３１】類似度の計算式の例について説明する。質問文とデータベース文ｐの類似度は，以下の式のScore(p)によって与えられる。

【００３２】$Score(p)=BNST1(p)+\alpha\times BNST2(p)-\beta_1\times BNUM_{sent}(p)-\beta_2\times BNUM_{all}(p)$

ここで，

$BNST1(s)=\Sigma\ NEAR(p,b)\times JIRITSU(b)$　（Σは質問文の文節ｂの和）

$BNST2(s)=\Sigma\ NEAR(p,b)\times bnst2(b1,b2)$　（Σは質問文のすべての係り受け関係(b1,b2) で和をとる。ただし，b1はb2に係る）

bnst2(b1,b2)は，JIRITSU(b1)×FUZOKU(b1)×JIRITSU(b2) が０でないとき，

bnst2(b1,b2)＝ JIRITSU(b1)＋FUZOKU(b1)＋JIRITSU(b2)

それ以外のとき，bnst2(b1,b2)＝０

NEAR(p,b) ＝NEAR1(p,b)＋NEAR2(p,b)＋NEAR3(p,b)

NEAR1(p,b)は，ｂが解答部分と同一文の場合，NEAR1(p,b)＝γ_1それ以外のとき，NEAR1(p,b)＝１

NEAR2(p,b)は，ｂが解答部分と同一文で疑問詞とｂの係り受け距離とデータベース文でそれらに対応する文節間の係り受け距離の大きいほうがｄの場合，

NEAR2(p,b)＝$1+\gamma_2$ ／（１＋ｄ）

それ以外のとき，NEAR2(p,b)＝１

NEAR3(p,b)は，ｂが解答部分と同一文で疑問詞とｂの文節距離とデータベース文でそれらに対応する文節間の文節距離の大きいほうがｄ′の場合，

NEAR3(p,b)＝$1+\gamma_3$ ／（１＋ｄ′）

それ以外のとき，NEAR3(p,b)＝１

抽出したデータベース文のすべての文節は，上記のScore(p)の値が最大になるように入力側のいずれかの文節に対応づける。JIRITSU(b)は，入力側の文節ｂと，それに対応づけられた文節との間の自立語における類似度で，FUZOKU(b) は，入力側の文節ｂと，それに対応づけられた文節との間の付属語における類似度である。

【００３３】$BNUM_{sent}$ (p) は，データベース文ｐのうち解答部分が含まれる文の文節数である。$BNUM_{all}$ (p) は，データベース文ｐの文節数である。二つのBNUMは，他の情報が同じなら余分な文節が存在しない文との照合のほうが大きくなるようにするための項である。

【００３４】NEARは，解答部分と近接している文節の値を上げるもので，質問文とデータベース文との照合をよりよく行うためのものである。NEARの算出で用いる二つの文節の間の「係り受け距離」とは，構文木におけるその二つの文節の間の枝の数を意味し，二文節間の「文節距離」とは，その二つの文節の間に１を加えた数を意味する。

【００３５】α，β_1，β_2，γ_1，γ_2，γ_3 は，実験で定める定数である。また，ここで示したScore(p)の式は，BNST1,BNST2 と二項関係までしか用いていないが，さらに三項，四項関係といったものを追加して用いてもよい。

【００３６】以上の式により質問文とデータベース文との類似度を計算したならば，この組合せのうち，最も大きい類似度の値をＳ，そのときの解の候補をＡとする（ステップＳ６）。

【００３７】次に，今回のＳの値が前回のＳよりも大きいかどうかを判定し（ステップＳ７），大きい場合にはステップＳ８へ進み，大きくない場合にはステップＳ９へ進む。

【００３８】ステップＳ８では，変形部１６が，変形規則記憶部１５に記憶されている変形規則を用いて，質問文の集合，データベース文の集合の各成員を書き換え，書き換えた文を質問文の集合，データベース文の集合にそれぞれ追加する。その後，ステップＳ６へ戻り，再度，照合部１４による照合を繰り返す。なお，変形部１６による文の変形処理を繰り返すと，質問文の集合，データベース文の集合の成員の数が膨大な数となり，計算コストが膨大になる。このときには，照合部１４による処理の際に，質問文とデータベース文の類似度の値がある程度大きくなる場合の成員のみを残して，それ以外の成員を削除しながら，処理を繰り返すのがよい。

【００３９】ステップＳ７の判定において，今回のＳの値が前回のＳよりも大きくならなかった場合，解答出力部１７は，最も類似度の大きい解の候補Ａから解を抽出し，それを質問文に対する応答文として出力し（ステップＳ９），処理を終了する。解の候補となったデータベース文からの応答文の生成は，例えば質問文における疑問詞を含む文節と対応づけられたデータベース文中の文節を解とすることで行う。

【００４０】以上の処理の例では，類似度の向上がみられなくなるまで変形部１６による変形を繰り返すとして説明したが，計算時間の関係上，類似度がある閾値を上回った場合に変形処理の繰り返しを打ち切ったり，または変形処理の繰り返しの最大回数をあらかじめ定めておき，その回数分繰り返した後に変形処理を終了したりするような実施も可能である。

【００４１】以下に，上記処理による具体的な実行例について説明する。

【００４２】(1) まず，質問文として以下の文が入力されたとする。

＊【００４３】

・質問文：「日本の首都はどこであるか」 ── a

この文が質問文の集合の成員となる。キーワード抽出／情報検索部１２では，「日本」「首都」がキーワードとして得られる。

【００４４】(2) キーワード抽出／情報検索部１２は，「日本」と「首都」をキーワードとしてデータベース１１を検索し，これらのＩＤＦの値が大きい文を複数文，抽出する。ここでは，以下の二つの文が得られたとする。

＊【００４５】

・データベース文１：「東京は日本の首都である」 ── b

・データベース文２：「日本の隣国韓国の首都はソウルである」 ── c

この二つの文が抽出した文の集合の成員となる。

【００４６】(3) 構文解析部１３は，質問文とデータベース文をすべて構文解析する。つまり，上記ａ，ｂ，ｃの文を構文解析する。

【００４７】(4) 照合部１４は，照合の開始にあたって，最大の類似度を記憶する変数Ｓの値を０にセットする。

【００４８】(5) その後，照合部１４は，質問文の集合の成員とデータベース文の集合の成員をあらゆる組合せで照合する。つまり，ａとｂの照合，ａとｃの照合を行う。ここで，ａとｂの類似度が２２，ａとｃの類似度が３５であったとする。このとき，照合の最大の類似度は※

※３５であるので，Ｓは３５にセットされる。

【００４９】(6) Ｓの値３５が前回のＳの値０に比べて大きいので，変形部１６の処理（ステップＳ８）が実行される。変形部１６で，質問文の集合，データベース文の集合の各成員に変形規則が適用される。ここでは，説明を簡単にするため，変形規則が以下のものだけ用いられていたとする。

【００５０】

・変形規則：ＡはＢである→ＢはＡである ─── d

上記ａ，ｂ，ｃの文にそれぞれｄの変形規則が適用され，以下の三つの文が新たに生成される。

【００５１】

・ａの変形（ｄの利用）：「どこが日本の首都であるか」 ─── e

・ｂの変形（ｄの利用）：「日本の首都は東京である」 ─── f

・ｃの変形（ｄの利用）：「ソウルは日本の隣国韓国の首都である」── g

これらを質問文の集合，データベース文の集合に追加する。ここでは，ｅが質問文の集合に，ｆ，ｇがデータベース文の集合に追加される。

【００５２】(7) 再度，照合部１４で質問文の集合の成員とデータベース文の集合の成員をあらゆる組合せで照合する。つまり，ａとｂ，ａとｃ，ａとｆ，ａとｇ，ｅとｂ，ｅとｃ，ｅとｆ，ｅとｆの照合を行う。ここでは，ａとｆ，ｅとｂの類似度が４７で最も大きかったとする。Ｓは，４７にセットされる。

【００５３】(8) 今回のＳの値４７は前回のSの値３５よりも大きいので，再度，変形部１６の処理が実行される。変形部１６では，今回はすでに質問文の集合，データベース文の集合にある文しか生成されないので，質問文の集合，データベース文の集合は変化しない。

【００５４】(9) 再度，照合部１４で質問文の集合の成員とデータベース文の集合の成員をあらゆる組合せで照合する。このとき，集合の成員が前回から変化していないので，前回と同じくａとｆ，ｅとｂの照合の類似度が最も大きく，その値は４７となる。また，このときの解の候補は照合の際に疑問詞の部分に対応していた「東京」であるとする。

【００５５】(10) 今回のＳの値４７は前回のＳの値と同じ大きさなので，繰り返し処理はここで終了し，解の候補としていた「東京」が解として出力される。

【００５６】本発明では，さらに変形部１６が使用する変形規則を自動獲得する変形規則自動獲得システム２を持つ。この変形規則自動獲得システム２は，質問応答システム１内の処理機能として，質問応答システム１内に組み込むこともできる。

【００５７】従来技術として説明したＫａｔｚの方法では，変形規則は，ある特定の基底表現に変形するものであるため，人手によって変形規則を記述し作成する必要がある。これに対し，本発明で用いる変形規則は，文Ａから文Ｂへの変形と，文Ｂから文Ａへの変形とを区別する必要がない。したがって，以下に説明するように，既存のデータを用いてコンピュータによる処理により，変形規則を自動生成することが可能である。

【００５８】第１の方法は，コンピュータが読み取り可能な複数の国語辞典を用意し，これら複数の辞典の説明文（定義文）の突き合わせにより，同義語・同義フレーズに関する知識を得て，それから変形規則を獲得する方法である。

【００５９】第２の方法は，質問応答集のデータを与えることで，その質問応答を成立させるために必要とされ

る同義語・同義フレーズに関する知識を得て，変形規則を獲得する方法である。

【００６０】変形規則を自動獲得する第１の方法の処理フローチャートを，図４（Ａ）に示す。まず，複数の国語辞典等の辞書ファイルを用意し，それらから読み出した辞書データから，同じ見出し語（単語）に対する説明文を抽出する（ステップＳ１１）。

【００６１】次に，抽出した複数の説明文を突き合わせ（ステップＳ１２），その結果から，一致する部分を除いた異なる表現の部分を，同義語もしくは同義フレーズとして抽出する（ステップＳ１３）。その同義語もしくは同義フレーズを相互に変換できるように変形規則の左辺，右辺に割り当てることにより，変形規則を生成し，記憶する（ステップＳ１４）。以上の処理を辞書データ中のすべての見出し語について繰り返し，すべての見出し語について処理したならば，処理を終了する（ステップＳ１５）。

【００６２】具体例で説明する。例えば辞書＃１と辞書＃２の二つの辞書があったとする。そこで「あべこべ」という見出し語の辞書データから変形規則を生成するものとする。

【００６３】辞書＃１の「あべこべ」の定義文が，「順序・位置などの関係がさかさまに入れかわっていること」であったとし，辞書＃２の「あべこべ」の定義文が，「順序・位置・関係がひっくり返っていること」であったとする。同じ見出し語の定義文であるので辞書＃１の定義文と辞書＃２の定義文とは，同じ意味であると考えられる。これらの定義文を突き合わせて照合すると，「順序・位置などの関係が」と「順序・位置・関係が」の部分がよく似ているので，この部分は一致すると考えられる。このことから，両定義文の残りの部分「さかさまに入れかわっていること」と「ひっくり返っていること」が対応することがわかる。これから，以下の二つの変形規則が得られる。

【００６４】変形規則１：「さかさまに入れかわっている」→「ひっくり返っている」
変形規則２：「ひっくり返っている」→「さかさまに入れかわっている」
ここでは，複数の国語辞典の定義文の対応関係を利用しているが，これ以外にも意味的な対応関係があるもの同士ならば，上記第１の方法を使うことができる。

【００６５】変形規則を自動獲得する第２の方法の処理フローチャートを，図４（Ｂ）に示す。まず，電子化された質問応答集のテキストデータから質問文と応答文とを読み出す（ステップＳ２１）。

【００６６】次に，読み出した質問文と応答文とを突き合わせ（ステップＳ２２），その結果から，同義語もしくは同義フレーズとして抽出する（ステップＳ２３）。その同義語もしくは同義フレーズを相互に変換できるように変形規則の左辺，右辺に割り当てることにより，変形規則を生成し，記憶する（ステップＳ２４）。以上の処理をすべての質問応答文について繰り返し，すべての質問応答文について処理したならば，処理を終了する（ステップＳ２５）。

【００６７】具体例で説明する。今，質問応答集から得た質問文と応答文として，次のような文があったとする。

【００６８】質問文：「日本の首都はどこであるか」
応答文：「東京は日本の首都である」
このとき，おおざっぱな照合でも，場所を意味する疑問詞「どこ」と「東京」とが対応することがわかる。また，「日本の首都」というフレーズは容易に対応していることがわかる。以上の知識から，以下の変形規則を得ることができる。なお，疑問助詞「か」は省略する。

【００６９】変形規則：「ＡはＢである」→「ＢはＡである」
この変形規則があると上記の例の場合，質問文とデータベース文とが完全に一致するまで変形することができる。このように，質問文とデータベース文の類似度が極力上がるような変形規則を，フレーズの対応関係などから獲得する。

【００７０】本発明の質問応答システムは，情報抽出の技術としても有用である。情報抽出とは，例えば県に関する情報であれば，県名，県庁所在地，面積，人口，主な産物…といった情報を，既存のデータベースから自動抽出する技術である。

【００７１】一般に，現在の情報抽出の技術は，対象とする分野固有の知識に依存する部分が多く，システムを他の分野へ移行させるのにコストがかかるという問題があるとされている。

【００７２】これに対し，本発明のような質問応答システムは，例えば「岩手県の県庁所在地はどこですか」と聞いて，「盛岡」と答えるシステムである。質問応答の場合，多様な自然言語の質問を行うことができ，分野依存性がなく，さまざまな情報を自由に取得できるという利点を持っている。この質問応答システムを用いて，前述した県に関する一覧的な情報を抽出するには，「県にはどのようなものがありますか」と聞いてから，それぞれに対して「県庁所在地はどこですか」，「面積はいくらですか」，…と順次質問していけばよい。この逐次的な質問をプログラム化しておけば，質問応答システムにより分野依存性なく情報抽出の問題を解くことができる。

【００７３】この結果，例えば次のような情報を自動抽出することが可能になる。

【００７４】
（青森県，青森，Ｘｋｍ^2，ａ人，りんご，……）
（秋田県，秋田，Ｙｋｍ^2，ｂ人，米，……）
（岩手県，盛岡，Ｚｋｍ^2，ｃ人，…，……）
・・・・・

【００７５】

【発明の効果】以上説明したように，本発明によれば，質問文とデータベース文の照合の際に，照合の類似度が上がるように質問文とデータベース文とを変形するので，質問文とデータベース文との照合の精度を向上させることができる。また，本発明では，類似度という尺度で変形操作を制御するので，変形規則として基底表現に限らず任意の変形規則を用いることができ，変形規則についても容易に記述または生成することが可能である。

【図面の簡単な説明】

【図１】本発明のシステム構成例を示す図である。

【図２】変形規則の例を示す図である。

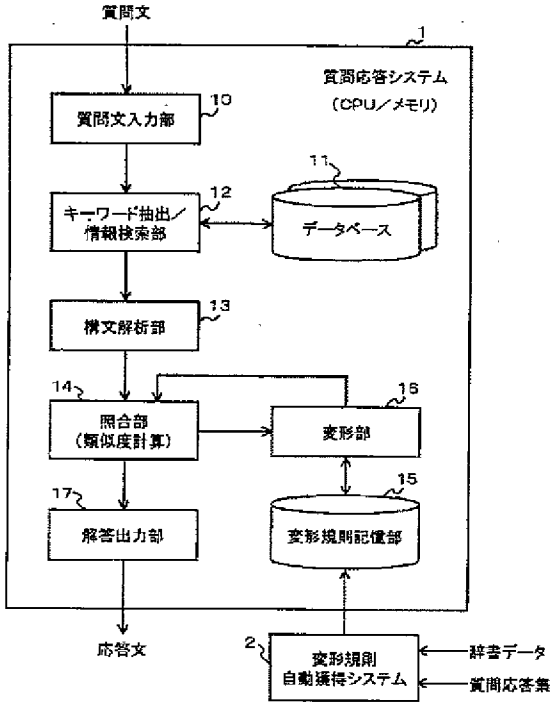
【図３】質問応答システムのフローチャートである。

＊【図４】変形規則自動獲得システムのフローチャートである。

【符号の説明】

１　質問応答システム
２　変形規則自動獲得システム
１０　質問文入力部
１１　データベース
１２　キーワード抽出／情報検索部
１３　構文解析部
１４　照合部
１５　変形規則記憶部
１６　変形部
１７　解答出力部

【図１】



【図２】

変　形　規　則　の　例

(A) 同義語の場合

アメリカ合衆国	→	米
米国	→	米
アメリカ	→	米
普通の	→	一般的な
人々	→	人たち
災害	→	被害

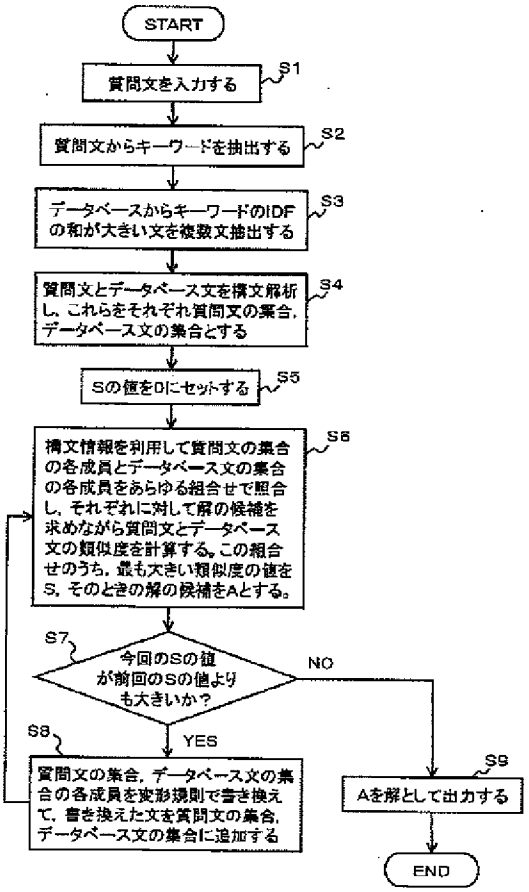
(B) 同義フレーズの場合

AはBである	→	BはAである
Xである	→	Xです
Xです	→	Xである
Xで構成されている	→	Xからできている
あります	→	ある
AはBである	→	BがAである

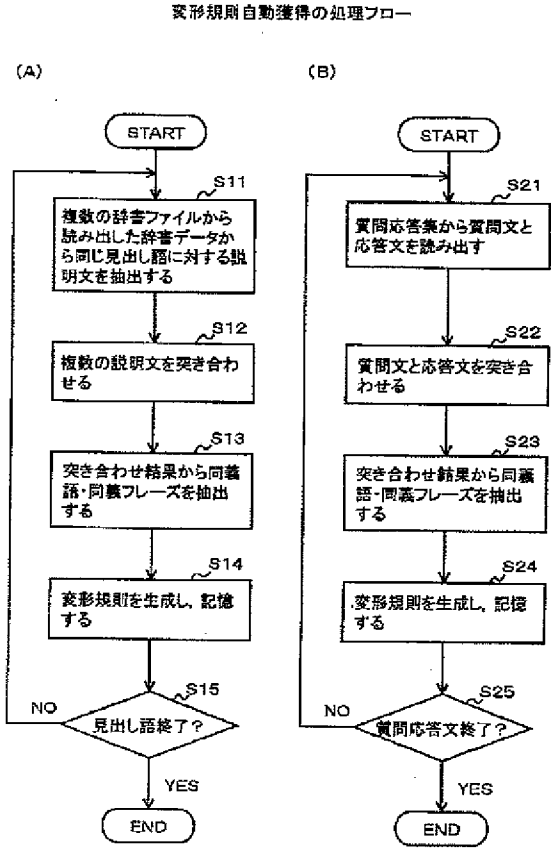
(C) 推論に関与する変形規則

(Aである)(AならばBである)	→	Bである

【図３】



【図４】



【手続補正書】
【提出日】平成１３年６月２１日（２００１．６．２１）
【手続補正１】
【補正対象書類名】明細書
【補正対象項目名】特許請求の範囲
【補正方法】変更
【補正内容】
【特許請求の範囲】
【請求項１】　自然言語による質問文を入力し，データベース中の文との照合によって応答文を生成して出力する質問応答システムにおいて，文を同じ内容を表す他の文に変形する規則を記憶する変形規則記憶部と，入力した質問文とデータベースから抽出した文とを照合し，それらの類似度を算出する照合部と，前記照合部による類似度の算出結果に基づき，前記変形規則記憶部に記憶されている変形規則を用いて前記質問文と前記データベースから抽出した文とを書き換える変形部と，前記照合部と前記変形部とによる処理を繰り返した後，前記類似度が最も高くなる照合において抽出された解を応答文として出力する解答出力部とを備えることを特徴とする質問応答システム。
【請求項２】　自然言語による質問文を入力し，データベース中の文との照合によって応答文を生成して出力する質問応答処理方法において，入力した質問文とデータベースから抽出した文とを照合し，それらの類似度を算出する過程と，あらかじめ記憶されている文の変形規則を用いて，前記質問文と前記データベースから抽出した文とを，それらの類似度が最も高くなるまで書き換える過程と，前記類似度が最も高くなる照合において抽出された解を応答文として出力する過程とを有することを特徴とする質問応答処理方法。
【請求項３】　自然言語で記述された文を同一言語により記述される同じ内容を表す他の文に変形する変形規則をコンピュータを用いて生成する方法であって，複数の同一言語により記述された意味的な対応関係がある言語情報を抽出する過程と，抽出した複数の言語情報を突き

合わせ，その結果から同義語または同義フレーズを抽出する過程と，抽出した同義語または同義フレーズから，ある文を同じ内容を表す他の文に書き換えるための変形規則を生成する過程とを有することを特徴とする変形規則自動獲得処理方法。

【請求項４】　自然言語で記述された文を同じ内容を表す他の文に変形する変形規則をコンピュータを用いて生成する方法であって，質問文とそれに対する応答文とを入力する過程と，入力した質問文と応答文とを突き合わせ，その結果から同義語または同義フレーズを抽出する過程と，抽出した同義語または同義フレーズから，ある文を同じ内容を表す他の文に書き換えるための変形規則を生成する過程とを有することを特徴とする変形規則自動獲得処理方法。

【請求項５】　自然言語による質問文を入力し，データベース中の文との照合によって応答文を生成して出力するためのプログラムを記録した記録媒体であって，入力した質問文とデータベースから抽出した文とを照合し，それらの類似度を算出する処理と，あらかじめ記憶されている文の変形規則を用いて，前記質問文と前記データベースから抽出した文とを，それらの類似度が最も高くなるまで書き換える処理と，前記類似度が最も高くなる照合において抽出された解を応答文として出力する処理とを，コンピュータに実行させるためのプログラムを記録したことを特徴とする質問応答処理プログラム記録媒＊体。

【請求項６】　自然言語で記述された文を同一言語により記述される同じ内容を表す他の文に変形する変形規則をコンピュータを用いて生成するためのプログラムを記録した記録媒体であって，複数の同一言語により記述された意味的な対応関係がある言語情報を抽出する処理と，抽出した複数の言語情報を突き合わせ，その結果から同義語または同義フレーズを抽出する処理と，抽出した同義語または同義フレーズから，ある文を同じ内容を表す他の文に書き換えるための変形規則を生成する処理とを，コンピュータに実行させるためのプログラムを記録したことを特徴とする変形規則自動獲得処理プログラム記録媒体。

【請求項７】　自然言語で記述された文を同じ内容を表す他の文に変形する変形規則をコンピュータを用いて生成するためのプログラムを記録した記録媒体であって，質問文とそれに対する応答文とを入力する処理と，入力した質問文と応答文とを突き合わせ，その結果から同義語または同義フレーズを抽出する処理と，抽出した同義語または同義フレーズから，ある文を同じ内容を表す他の文に書き換えるための変形規則を生成する処理とを，コンピュータに実行させるためのプログラムを記録したことを特徴とする変形規則自動獲得処理プログラム記録媒体。

フロントページの続き

(72)発明者　井佐原　均
兵庫県神戸市西区岩岡町岩岡588－２　郵政省通信研合研究所　関西先端研究センター内

Ｆターム(参考)　5B075 ND03 NK02 NK31 NK35 PP24 PR06 QM08 QP03